

June 5, 2003

Joseph J. Merenda
Director, Office of Science Policy and Coordination
Office of Pollution Prevention and Toxic Substances
USEPA Headquarters, 7201
Ariel Rios Building
1200 Pennsylvania Avenue, N. W.
Washington, DC 20460

Re: Docket Control Number OPPTS-2003-0016.
Issues Pertaining to the EPA's EDMVS -Mammalian Multigeneration
Reproduction Study Design

Dear Dr. Merenda

The American Chemistry Council (ACC or the "Council") has played an active role in the development and implementation of the endocrine disruptor screening and testing program (EDSP) for several years.¹ The Council supports the Agency's establishment of the Endocrine Disruptor Methods Validation Subcommittee (EDMVS) to provide technical advice and recommendations to EPA concerning the validation of endocrine disruptor screening and testing methods. ACC looks forward to the timely development and implementation of a scientifically sound EDSP.

The Council represents more than 90 percent of the productive capacity for basic industrial chemicals within the United States and its members are the leading companies engaged in the business of chemistry. EPA's endocrine disruptor screening and testing program (EDSP) may significantly affect the Council and its members. For that reason, the Council and its members have attempted to assist the Agency in developing and implementing its EDSP. In that

¹ The American Chemistry Council represents the leading companies engaged in the business of chemistry. ACC members apply the science of chemistry to make innovative products and services that make people's lives better, healthier and safer. ACC is committed to improved environmental, health and safety performance through Responsible Care, common sense advocacy designed to address major public policy issues, and health and environmental research and product testing. The business of chemistry is a \$460 billion enterprise and a key element of the nation's economy. It is the nation's largest exporter, accounting for ten cents out of every dollar in U.S. exports. Chemistry companies invest more in research and development than any other business sector. Safety and security have always been primary concerns of ACC members, and they have intensified their efforts, working closely with government agencies to improve security and to defend against any threat to the nation's critical infrastructure.



Responsible Care®

regard, ACC and its members actively participated in EDSTAC and are actively participating in the EDMVS.

The Council would like to bring to your attention the following issues pertaining to the mammalian multigeneration reproduction study design. It is only very recently that extensive revisions of the mammalian multigeneration reproduction study came into effect (EPA-1998, OECD-2001, Japan-2001). These revisions were implemented to specifically enhance sensitivity to detect adverse effects caused by substances operating through an endocrine mode of action (e.g., by adding endpoints such as sperm parameters, estrous cycling, developmental markers, more extensive parental histopathology, brain, spleen and thymus weights of weanlings, oocyte counting). It took global regulatory agencies 5-8 years to revise and internationally harmonize these guidelines. Further design changes would also require a coordinated approach with OECD to achieve international harmonization to ensure mutual acceptance of data.

The Council submitted extensive comments to EPA on December 14, 2001 concerning misperceptions on the part of some as to the sensitivity of existing internationally harmonized and standardized mammalian multigeneration reproduction study protocol. In the comments and analysis submitted at that time, we showed that the current multigeneration assay is sufficient to detect and characterize adverse effects of substances operating through an endocrine mode of action. In our comments of December 2001, we recommended that EPA should, prior to initiating laboratory studies of the multigeneration mammalian reproduction study design 1) conduct a detailed review of the available scientific literature, including, where applicable, studies available to EPA from pesticide registrations, to critically evaluate study designs and outcomes to determine if there are any actual data to support concern for study design questions pertaining to sample size and number/age and types of observations/measurements and 2) development of hypothesis and draft study protocol, using an appropriate number of substances, an appropriate route of administration and dosing regimen, to investigate the hypothesis.

The current EPA EDMVS reports² address some of the recommendations, but this effort falls well short in several critical areas. The Council is concerned, in particular, that EPA appears to have initiated laboratory studies without conducting a thorough examination of the available data and clear articulation of the appropriate study hypothesis. Attachment 1 describes in detail the Council's analysis and recommendations concerning the questions and proposals EPA has put forward for review and discussion by the EDMVS.

The information and study results provided in the reports² are informative even if they fall short in addressing some critical matters. As indicated by the attached analysis (Attachment 1), the study results in fact show that the existing test guideline is sensitive to, and does detect effects of, endocrine active substances. Therefore, EPA should not undertake changes to the existing protocol at this time. EPA should not initiate changes to such an expensive, labor

² Mammalian 2-Generation Assay Validation: History, Plan, and Questions for EDMVS June 6, 2003 Meeting and Report on the One-Generation Extension Study of Vinclozolin and Di-N-Butyl Phthalate Administered by Gavage on Gestational Day 6 to Postnatal Day 20 in CD (Sprague-Dawley) Rats

intensive and animal intensive study design until the Agency can clearly demonstrate that the proposed additions are necessary, effective and that they do not create technical logistical barriers which compromise the overall conduct and outcome of the study. In short, the Council believes EPA has the responsibility to demonstrate the added value and need of the information to be obtained from the proposed additions with respect to use of the information in the overall risk assessment process. Prior to implementing modifications to the study design, EPA must make a convincing case through additional validation studies that each of the proposed additions has value in terms of either increasing the sensitivity of the study design (i.e. shows adverse effects at lower dose levels than the existing protocol) or qualitatively enhances the ability to detect adverse health effects of endocrine active substances (i.e., identifies a substance as producing adverse effects via an endocrine mode of action that would not be so identified with the existing protocol). As explained in the attached analysis, EPA is proposing a number of changes that build redundancy upon redundancy, without adding additional value.

Since the ultimate question is whether or not a revised mammalian multigeneration rat reproduction study (with pup retention) is more effective for risk assessment purposes than the current OPPTS 870.3800 multi-generation test, the most direct and efficient way to address this question would be a side-by-side comparison of the current multigeneration protocol vs. an otherwise identical protocol except for pup retention. Compounds that have a recent, well-conducted multi-generation study that used the new guideline should be selected. Routes of exposure, dose levels, and end points need to be identical in order for a meaningful comparison to be made.

The Council appreciates this opportunity to provide early input on matters related to the EDMVS. We look forward to continuing our work with EPA and other interested parties on the validation of EPA's EDSP. Please don't hesitate to call me if you have questions.

Sincerely,

Original Signed By

Richard A. Becker, Ph.D., DABT
Public Health Team

Attachment 1. Comments on EPA's Proposal to Modify the Multigeneration Rat
Reproduction Study Design

cc. Jim Kariya, Office of Science Policy and Coordination, EPA
Gary Timm, Office of Science Policy and Coordination, EPA
Jane Smith, Office of Science Policy and Coordination, EPA

Attachment 1. Comments on EPA's Proposal to Modify the Multigeneration Rat Reproduction Study Design

A. General Comments

In the Council's December 14, 2001 comments to EPA, the Council provided a detailed analysis and critique of EPA's presumed assumption that the existing mammalian multigeneration assay may lack the sensitivity and power to detect and characterize certain types of endocrine-active chemicals, i.e., those that impair the development of the male reproductive tract. We will not repeat the critique here – the points still stand. EPA is reminded that the multigeneration mammalian reproduction study design recently underwent the extensive revision (EPA-1998, OECD-2001, Japan-2001). These revisions were specifically implemented to enhance sensitivity to detect adverse effects caused by substances operating through an endocrine mode of action (e.g., endpoints such as sperm parameters, estrous cycling, developmental markers, more extensive parental histopathology, brain, spleen and thymus weights of weanlings, oocyte counting). These enhancements improve the ability of this assay to detect endocrine-mediated effects. It took global regulatory agencies at least 5-8 years to revise and internationally harmonize these guidelines. Calling for further design changes at this time without solid scientific evidence to support such changes is counterproductive to all of the efforts that went into revising and harmonizing these guidelines on an international level.

Statements or perceptions by some that the multi-generation study has "severe limitations, particularly with regard to the detection of low incidence phenomena (e.g., reproductive tract malformations)" are simply incorrect and are not supported by the available data. For example, male reproductive tract malformations were not 'missed' in previous studies. Male reproductive tract malformations *were* found in the 'older' multigeneration study of linuron (summarized in the publicly available EPA Registration Eligibility Document (see ACC's Comments to EDPA of December 14 2001). In fact, these multi-generation studies were conducted according to pre-1998 guidelines, so a current two generation study would only be more effective at detecting such effects.

Furthermore, gross necropsy at weaning (pnd 21) can and does detect morphological differences. For example hypospadias, retained nipples/areolae, missing epididymis, missing / small testes and altered anogenital distance can be detected in weanlings, albeit the size and stage of the development of some structures require focused and skilled evaluation. For example, in pnd 21 evaluation for hypospadias requires focused attention because the prepuce has not yet separated from the penis. In fact all of these 'endpoints' were detected at pnd 21 in the RTI Study sponsored by EPA (Report on the One-Generation Extension Study of Vinclozolin and Di-N-Butyl Phthalate Administered by Gavage on Gestational Day 6 to Postnatal Day 20 in CD (Sprague-Dawley) Rats: retained nipples/areolae page 38, epididymis missing/small page 39, hypospadias page 39, altered AG distance page 38).

There is an overriding need for a single globally harmonized multi-generation protocol. Otherwise there is great potential for needless repetition of an extremely resource-intensive study (e.g., a two-generation study uses >3000 animals per compound) with little to no public health benefit. In the absence of a globally harmonized protocol, results will always be open to

potential criticism that the wrong protocol was followed. Similarly, the risk assessments based on these results will be similarly open to unwarranted criticism. The need for global harmonization of protocols is evident from the considerable efforts of EPA, OECD, Japan and other regulatory agencies to harmonize a wide range of test protocols.

In many facilities, increasing the numbers of animals (hence animal cages) would likely require that the study be housed in more than one animal room. Based on the RTI protocol, approximately 400 additional cages (based on an average of 4 retained males/litter x 100 litters) would be needed to house the extra F1 males for 10 weeks. This would be in addition to the approximately 240 cages for the F1 animals selected to breed the F2 generation. From a logistical perspective, such an increase not only would lead to higher costs, but also to impacts on the study design (more than 1 animal room) and study management. A study of this magnitude is likely to stress the capabilities of many testing laboratories, and would increase the probability of errors simply due to logistical complexity. Such concerns emphasize the need for adequate validation before implementation of new or substantially revised test guidelines.

B. Specific Comments on EPA's June 2003 Validation Plan

EPA should be commended for recognizing the status of the 2-generation guideline study as a globally accepted study that is considered the definitive test for evaluating potential reproductive toxicity. The 2-generation study has a long history of successful use, and with the many new end points added to this guideline in the last few years, it is even more robust than ever. End points such as sperm parameters, estrous cycling, anogenital distance, puberty markers, more extensive parental histopathology, brain, spleen and thymus weights of weanlings, and oocyte counting were added to address concerns about endocrine disruption. It took global regulatory agencies at least 5-8 years to revise and internationally harmonize these guidelines, and the effort of contract and industrial toxicology labs to add these end points has been considerable. As such, the consideration of *additional* end points above and beyond the new guideline should be approached very cautiously.

1.Does the EDMVS agree that the additional endpoints/clarifications proposed for the 2-generation assay (Table 1) are well characterized and that further validation of this set of endpoints for use in EDSP Tier 2 is unnecessary?

The EPA should be applauded for their decision to exclude the addition of other endpoints discussed during EDSTAC. The inclusion of neurobehavioral endpoints to the multi-generation study would make an already complex study unwieldy.

- a. *Anogenital distance in all animals in both F1 and F2 at birth (pnd 2).* The measurement of AGD is included in the current multi-generation study guidelines (OPPTS 870.3800); consequently, this endpoint would be considered validated. However, is there data to support the value of including AGD measurements for both the F1 and F2 offspring? As Clark (1998) states "Anogenital distance is such a sensitive measure of antiandrogenic activity that it has been used as the endpoint in pharmacological assays for antiandrogenic activity.

Conversely, an agent that affects anogenital distance likely acts through a mechanism involving androgen activity at some level. Nevertheless, this does not mean that anogenital distance should be used as a surrogate marker for subsequent endpoints because the dose that affects anogenital distance is not a reliable predictor of the dose at which effects on these other endpoints will occur.” Clark continues, stating, “slight-to-moderate effects on anogenital distance in rodents that occur without other morphological effects do not affect function” (Clark, 1998). Thus, it seems prudent to consider use of AGD as a triggered endpoint. In this way, alterations in AGD would be linked to changes in other endocrine-mediated endpoints. Furthermore, AGD lacks some specificity as it can be altered by factors not specific to exogenous agents (e.g., prenatal uterine position, stress, altered arachidonic acid cascade, etc.), as well as factors that influence pup body weight (maternal or neonatal toxicity, litter size, etc.) (see Gallavan *et al.*, 1999). By retaining AGD as a triggered endpoint, these interpretive problems may be avoided.

- b. *Areola/nipples: what, where, how many, in both males and females, F1 and F2, pnd 13. Also at necropsy for F1 only.* While the assertion that animals are examined macroscopically for structural abnormalities is correct, this is not equivalent to recording areola/nipple number and location for all F1 and F2 offspring on pnd 13 and adult F1 necropsies. While the EPA recognizes the addition of a new time point for examination (pnd 13; shaving is not required at this stage of development) and the inclusion of larger numbers of animals (all vs. 3/sex/litter), it failed to note the additional effort needed to record these data, including the necessity to shave all 240 animals at necropsy (this assumes 30 males and 30 females per dose level). There are real concerns that such proposed changes would not be practical, given the labor intensiveness of the existing activities required of the technicians. In addition, proficiency would need to be established by each prosector in order to maintain variability to an acceptable degree.

While the additional work is not a problem in itself, the added value of this measure in view the additional effort is questionable. In a brief perusal of the literature, areola/nipple retention is seldom the most sensitive measure of altered endocrine status. In fact, more traditional endpoints already included in the multi-generation study are more likely to detect these changes, as indicated in Table 1.

In one study (Turner *et al.*, 2002), changes in nipple retention were not permanent, which raises another question as to whether measurements in the adults are needed. A more thorough review of the literature is needed to determine if areola/nipple retention is a sensitive and specific measure to detect altered endocrine status. Furthermore, areola/nipple retention, while used in investigative studies focused on anti-androgens, has not been validated across multiple laboratories.

Table 1. Comparison of retained nipples/areolas to other endpoints

| Reference | Compound and doses | Lowest dose showing retained areolas/nipples | Other altered endocrine/reproductive endpoints |
|-------------------------------|-------------------------------------------|----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| McIntyre <i>et al.</i> , 2002 | Linuron: 50 mg/kg/day | 50 mg/kg/day at pnd 13, 35 and 56 | Hypospadias, cleft penis, decreased testis and epididymis weights (for organs with normal appearance); altered testicular histology decreased AGD at pnd 1 and 56 |
| McIntyre <i>et al.</i> , 2001 | Flutamide: 6.25, 12.5, 25 and 50 | ≥ 6.25 mg/kg/day | At ≥ 6.25 mg/kg/day: increased cryptorchid/ectopic testes; cleft prepuce; decreased organ weights (epididymides, seminal vesicles, ventral and dorsolateral prostate); decreased AGD at pnd 1 and 100;. |
| McIntyre <i>et al.</i> , 2000 | Linuron: 12.5, 25, 50 mg/kg/day | 50 mg/kg/day | At 50 mg/kg/day: decreased pup survival to weaning; At ≥ 25 mg/kg/day: testicular hypoplasia; epididymal hypoplasia; altered testicular histology (some minimal indications of these effects at 12.5 mg/kg/day) |
| Turner <i>et al.</i> , 2002 | Fenitrothion: 5, 10, 15, 20, 25 mg/kg/day | 25 mg/kg/day; transient effect | At 20 and 25 mg/kg/day: Maternal toxicity; increased fetal death; At 25 mg/kg/day: decreased AGD (transient effect) |

- c. *TSH, T4, thyroid weight, and thyroid histology, all at necropsy.* Thyroid endpoints are not included in the current multi-generation study guideline and appear to be useful additions to consider for the detection of thyroid-active agents. Also, we are in agreement with the EPA not to include T3 as a required endpoint. However, we feel strongly that thyroid endpoints should be evaluated in adult animals. While the EPA has proposed measuring T4 and TSH

levels on pnd 21, work by Döhler *et al.* (1979) has documented that the levels of serum T4 change with age in the rat prior to adulthood. T4 peaks at approximately day 15 in young rats, then declines until approximately pnd 27, at which time the serum levels begin to rise again. T4 also increases markedly in male rats from pnd 33 to 50. With such dynamic changes in T4 levels, the sensitivity of this measure to detect true changes in thyroid activity will be greatly diminished. Accurate sampling of serum T4 and TSH levels is already complicated by the sensitivity of this measure to stress and time of day of sampling; thus, it seems unwise to further confound these endpoints with age-related effects.

Further, there is some evidence to suggest that thyroid hormone concentrations may be altered in females with stages of the estrous cycle. Differing T4 levels have been reported between the nadir during metestrus/diestrus and the peak during proestrus (Döhler *et al.*, 1979). As rats age, female rats are more efficient at hepatic T4 to T3 deiodination (da Costa *et al.*, 2001). This, coupled with the slightly greater sensitivity of male rats to thyroid perturbations, raises the question as to whether an evaluation of thyroid function in male rats alone may be sufficient. In this light, EPA should address the question as to whether thyroid hormone analyses in females is necessary.

The critical question that EPA needs to address through validation is whether the proposed additional measurements of T4 and TSH are relevant and reliable for evaluating thyroid function in the multigeneration study.

- d. *Whole-mount histology of mammary tissue in males, triggered if abnormalities are seen in gross examination.* The current guidelines state that histological examination of treatment-related abnormalities should be considered if such evaluation were deemed appropriate and would contribute to the interpretation of the study data. Presumably, the EPA's addition is intended to require whole mount histological examination of mammary tissue in males with retained areola/nipples (presumably, whole mount refers to compression of mammary glands prior to embedding and sectioning). It is difficult to discern how histological examination of these tissues will provide additional pertinent information when persistence of these structures in adult males is itself an indication of altered development. As proposed, EPA has not provided scientific justification for such an expanded effort, and thus ACC cannot support this proposed addition. At the very least, EPA needs to clarify what is proposed here.
- e. *Testis location at necropsy (descended/undescended, attached/floating).* Testis location is already examined during gross necropsy of weanling and adult males. If a deviation from the controls is noted, this information is recorded.
- f. *Malformation, agenesis, or inappropriate presence of any of the sex organs (e.g., prostate agenesis, presence of uterus in male).* Reproductive organs are the primary focus of gross pathological examinations for weanling and adult animals. Once again, deviations from the controls are noted and this information is recorded.

- g. *Number of days until the plug is observed should be analyzed as an indirect indicator of sexual behavior.* Laboratories already record the day on which females mate (i.e., have vaginal plugs *in situ* or vaginal smears positive for the presence of sperm) and most laboratories report “days to mating”. Addition of this parameter as a reporting requirement should not pose difficulties.
- h. *Prostate weight by lobe (ventral and dorsolateral).* This addition proposes adding ventral and dorsolateral prostate weights to the current requirement to measure whole prostate weight. Once again, the value of these additional endpoints should be evaluated through validation studies, particularly as to whether separate measurement of ventral and dorsolateral prostate weights would enhance the specificity and sensitivity of the assay in comparison with whole prostate weight and histology.

EPA must document with scientific data that there is additional information that is critical to have that can only be obtained by dissecting the prostate and weighing each lobe separately. If this proposal is to be evaluated, then additional data is needed as part of EPA’s validation efforts.

Experience with the Hershberger assay in the OECD validation program failed to show any meaningful difference in overall response when a comparison was made between fresh weight and fixed weight of the ventral prostate. A similar comparison needs to be made by EPA for the proposal to obtain separate ventral and dorsal lobe prostate weights.

- i. *Histology on testis/ovary for 2 females and 2 males per litter of the F1 generation (that is, the pair used for breeding the F2, and one additional pair) (= 40 animals per sex, assuming 20 litters).* While the EPA requires histopathology for “ten randomly chosen high dose and control P and F1 animals per sex”; the OECD and JMAFF guidelines (OECD 416) require that “full histopathology of the preserved organs and tissues (listed in paragraph 42) should be performed for all high dose and control P and F1 animals selected for mating.” Note that testis and ovaries are included in paragraph 42. Because most laboratories wish to conduct multi-generation studies that will be accepted internationally, most laboratories would follow the more stringent guidelines and examine all animals in accordance with OECD requirements. Thus, laboratories are already examining 20-30 animals per sex per dose level. Thus, the addition of a second animal from each litter is unnecessary.

Oocyte quantitation methodology was reviewed by a recent ILSI workshop (ILSI, 1998), which included an extensive discussion of power analyses for this end point. A general recommendation was that a sample size of 10/dose group would provide adequate power when this end point was added to a multigeneration study. Therefore, the proposal to sample two F1 animals/litter from at least 20 litters/dose group is not warranted. This and other critical methodological issues identified at the ILSI workshop should be addressed as part of a validation exercise prior to considering inclusion of any expanded effort into routine testing. Furthermore, as part of an effort to establish relevancy & reliability, the degree of change associated with 'adversity' needs to be addressed. The experts participating in the

ILSI workshop were unable to reach consensus regarding what degree of change would be considered adverse.

2. Does the EDMVS agree that the endpoints in the Tier 2 assay (including the endpoints proposed in Table 1) will allow a compound to be identified as possibly having “an effect in humans that is similar to an effect produced by a naturally occurring estrogen” (or androgen/anti-androgen or thyroid mimic/inhibitor) *in the absence of Tier 1 data*? (That is, for chemicals which voluntarily bypass Tier 1, will a Tier 2 assay that includes these endpoints allow identification of a chemical as meeting the requirements of FQPA?) If not, what other endpoints should be included, or what supplemental testing would be appropriate?

The Council believes that the existing multigeneration study design is sensitive and effectively detects potential endocrine mediated adverse effects, including substances that act via estrogen, androgen or thyroid hormone modes of action. There is an extensive database which shows that this study design is robust and useful for human health risk assessment purposes. Contrary to some misstatements or incorrect perceptions, the multi-generation study does not have “severe limitations, particularly with regard to the detection of low incidence phenomena (e.g., reproductive tract malformations). See ACC’s comments to EPA of December 14, 2001.

Furthermore, the mammalian multigeneration reproduction study design was very recently revised (EPA-1998, OECD-2001, Japan-2001). These revisions were implemented to specifically enhance sensitivity to detect adverse effects caused by substances acting through an endocrine mode of action (e.g., by adding endpoints such as sperm parameters, estrous cycling, developmental markers, more extensive parental histopathology, brain, spleen and thymus weights of weanlings, oocyte counting etc.). It took global regulatory agencies 5-8 years to revise and internationally harmonize these guidelines. Further design changes would also require a coordinated approach with OECD to achieve international harmonization to ensure mutual acceptance of data.

If EPA is to undertake additional validation efforts, then EPA has the responsibility to demonstrate the added value and need of the information to be obtained from the proposed additions with respect to use of the information in the overall risk assessment process. In short, prior to implementing modifications to the study design, EPA must make a convincing case that each and every one of the proposed additions has value in terms of increasing the sensitivity of the study design and enhancing the ability to detect adverse health effects of endocrine active substances. As explained in this analysis, EPA appears to be proposing a number of changes that build redundancy upon redundancy, without adding additional value.

3. Does the EDMVS agree that the procedures and endpoints in Table 2 should be listed explicitly (even though already covered in the Guideline), to ensure adequate examination?

Contrary to the statements made in EPA’s first paragraph, the components in the list are not necessarily “specific endpoints already covered by the current Reproductive Toxicity Guideline”. That is, all endpoints listed in Table 2 are not specifically examined during a multi-generation

study, nor are they required to be evaluated in the current version of the guidelines (OPPTS 870.3800).

Furthermore, a number of the additional measurements/endpoints/observations are unnecessarily redundant. EPA needs to demonstrate that any new discrete measurement is appropriate, reliable and adds value to the overall study outcome. Simply adding measurements or observations without justification is not appropriate.

Males – Necropsy after Puberty

1. Body weight and gross anomalies are recorded at necropsy.
2. At this time, laboratories are not shaving the ventral surface of all males from the inguinal region to the neck to count and record the position of areolas/nipples. This is not required in the current guidelines, nor is it necessarily “value added” (see discussion above).
3. While males are examined for hypospadias, epispadias and cleft phallus (these effects would be readily apparent during preputial examinations), AGD is not measured in adult males.
4. As mentioned, undescended testes would be readily apparent at necropsy (if not recorded during the in-life phase of the study).
5. Perineal soiling is recorded at necropsy.
6. Each animal is examined for the age at which preputial separation occurs. If a preputial thread persists, laboratories would record this. These observations are typically recorded in conjunction with a preputial examination rather than at adult necropsy.
7. For animals on study past puberty, the age and body weight at preputial separation is recorded. Age at the onset of preputial separation is not typically recorded, but rather, most labs report the day on which preputial separation is complete. Body weight is collected at the time of completion.
8. Most laboratories record the body weights of animals once each week after weaning, not twice per week as indicated here.

Internal Endpoints

9. If the location of a testis is abnormal, this finding would be recorded at necropsy.
10. As mentioned above, if the testis is located in an abnormal position, this finding would be recorded. The presence or absence of the gubernacular cords would likely not be recorded and the length of the gubernacular cords would not be measured. It is sufficient to identify that the testis failed to reach the appropriate site. The identification of the structural reason for this failure is not necessary during hazard identification and characterization.
11. The presence of the cranial suspensory ligaments is not typically evaluated at necropsy. However, the likelihood that the cranial suspensory ligaments would be present in the males in absence of other detectable effects (e.g., ectopic gonad) seems remote.
12. If testes are small, absent, fluid-filled, enlarged, appear infected, etc., this finding would be recorded at necropsy.
13. Similarly, gross abnormalities in the epididymides would be recorded.
14. If the ventral prostate is noted to be small during gross examination, this finding would be recorded as decreased prostate size. The guidelines do not require that the specific lobes be identified.

15. If the dorsolateral prostate is noted to be small during gross examination, this finding would be recorded as decreased prostate size. The guidelines do not require that the specific lobes be identified.
16. If the seminal vesicles are grossly abnormal, this would be recorded at necropsy.
17. If the coagulating glands are noted to be increased or decreased in size during gross examination, this finding would be recorded.
18. If the kidneys display hydronephrosis and/or calcium deposits, this would be recorded at necropsy.
19. Hydroureter would be recorded at necropsy.
20. The presence of bladder stones or blood in the bladder would be noted at necropsy.

Weights and Histology

21. The guidelines do not require that each testis be weighed separately. The guidelines require merely that the testes are weighed. Most labs weigh one testis separately because of the need to determine tissue efficiency subsequent to spermatid counts. Often, however, the reproductive study results present and statistically analyze a paired testes weight.
22. Again, the guidelines require that epididymides and at least one cauda epididymis are weighed; they do not specify that the various components of the epididymides must weighed (caput, corpus and cauda). Again, many labs present a paired epididymal weight, then record a cauda epididymis weight for the determination of tissue efficiency after sperm counts have been completed.
23. See comment for 22.
24. Seminal vesicle weights with coagulating glands are currently measured according to the guidelines. Histology also is examined.
25. Currently, labs are required to collect prostate weights (not ventral and dorsolateral prostate weights). Histology is conducted on the whole prostate, which includes an examination of both the ventral and dorsolateral sections.
26. While laboratories are required to collect kidney weights, these may be collected as paired weights. The guidelines do not require these weights to be collected separately. Histological examination of the kidneys is not required.
27. Paired adrenal weights are collected in accordance with the guidelines and these glands are examined histologically.
28. Liver weight is collected according to the guidelines, but histological examination of the liver is not required..
29. Evaluation of the weight and histology of the levator ani-bulbocavernosus muscle is not required in the current guidelines.
30. Cowper's gland weights and histology are not required in the current guidelines.
31. The glans penis weights and histology are not required in the current guidelines.
32. See comments in number 25.
33. While brain weight is required, brain histology is not required in the current guidelines.
34. The pituitary is weighed and examined histologically according to the current guidelines.
35. Thyroid weights and histology are not required in the current guidelines. However, ACC recognizes the value of these endpoints in assessing thyroid function. ACC favors inclusion of these endpoints in order to identify potential thyroid-active agents.
36. Neither heart weights nor histology are required in the current guidelines. Is the EPA's suggestion to include heart weight and histology if a material is suspected of being an

antithyroid agent? If so, the specific thyroid endpoints are examined post-necropsy (i.e., fixed thyroid weight, histology, serum TSH and serum T4), thus, what endpoints would trigger the collection of heart weights and histology?

Females – Necropsy

37. Body weight and gross anomalies are recorded at necropsy.
38. At this time, laboratories are not shaving the ventral surface of all females from the inguinal region to the neck to count and record the position of areolas/nipples. This is not required in the current guidelines, nor is it necessarily “value added” (see discussion above).
39. While females are examined for hypospadias and cleft phallus (these effects would be readily apparent during vaginal opening examinations), AGD is not measured in adult females. Furthermore, ano-vaginal distance (AVD) is not typically measured in reproduction studies and is not required in the guidelines.
40. Each animal is examined for the age at which vaginal opening occurs. If a vaginal thread persists, laboratories record this. These observations are typically recorded in conjunction with vaginal opening examinations rather than at adult necropsy.
41. See comments for number 40.
42. If mammary tumors are present, this observation would be recorded at necropsy. Histological examination of mammary tissue is not specifically required by the guidelines; however, histological examination of gross lesions is required. Therefore, gross alterations in mammary tissue would be examined microscopically.

Internal Observations

43. Any abnormalities in the position, size or color of the ovaries would be recorded at necropsy.
44. The absence of the cranial suspensory ligaments would likely be noted at necropsy as an “ectopic ovary”. This specific endpoint (presence of the cranial suspensory ligaments) is not typically evaluated at necropsy nor is it required by the guidelines.
45. The presence of follicular cysts on an ovary or atrophy of an ovary would be noted at necropsy.
46. Absence of the lower vagina would be noted at necropsy.
47. Gross uterine abnormalities (e.g., bi- or unilateral agenesis of the oviducts of the uterine horns, infections, hydrometrocolpos, etc.) would be noted at necropsy.
48. The presence of any male tract tissues (e.g., ventral prostate, seminal vesicles, Cowper’s glands, levator ani-bulbocavernosus muscle) would be noted during necropsy. These tissues would be saved for histological examination as gross lesions.

Necropsy weights and Histology

49. Body weights and organ weights (liver, kidneys, adrenals, brain, and pituitary) are required according to the guidelines. According to the guidelines, the adrenals and pituitary are examined histologically, whereas the liver, kidneys and brain are not. As stated above, neither heart weights nor heart histology is required in the current guidelines. Is the EPA’s suggestion to include heart weight and histology if a material is suspected of being an antithyroid agent? If so, the specific thyroid endpoints are examined post-necropsy (i.e., fixed thyroid weight, histology, serum TSH and serum T4), thus, what endpoints would trigger the collection of heart weights and histology?

50. The ovaries are weighed and examined histologically. Primordial follicle counts also are conducted on the P2 (F1) females.
51. The oviducts are evaluated histologically, usually in conjunction with the uterus.

4. If EDMVS advises EPA to validate additional endpoints,

- a. can “new” endpoints be validated separately from endpoints already in the reproductive toxicity assay? (I.e., is it scientifically acceptable to examine the relevance and reliability of *endpoints* or must we validate the entire *assay*?)**
- b. is it necessary to validate all new endpoints in a 2-generation study, or can relevance and reliability be established in a shorter assay, such as a one generation protocol or an in-utero-through-lactation protocol?**
- c. how many laboratories should be required for interlaboratory comparability?**
- d. how many chemicals per mode of endocrine activity should be tested in validation? (e.g., ER/AR binding, each step of steroidogenesis, thyroid hormone transport protein binding, thyroid hormone metabolism, etc.)**

As discussed above, and based on our evaluations, the scientific basis for considering changes to the multi-generation study is lacking. However, if such changes must be considered, then the design changes of interest (i.e., retaining extra pups to adulthood) must be subjected to experimental validation before being considered for possible alteration of existing guidelines. What is essential for validation? Studies must be done which demonstrate that retaining pups truly changes the potential risk assessment, i.e., do such changes yield different and lower NOELs or lead to the identification of new target organs?

Since the ultimate question is whether or not a revised mammalian multigeneration rat reproduction study (with pup retention) is more effective for risk assessment purposes than the current OPPTS 870.3800 multi-generation test, the most direct and efficient way to address this question would be a side-by-side comparison of the current multigeneration protocol vs. an otherwise identical protocol except for pup retention. Compounds that have a recent, well-conducted multi-generation study that used the new guideline should be selected. Routes of exposure, dose levels, and end points need to be identical in order for a meaningful comparison to be made.

5. Does the EDMVS agree that the one-generation extension study shows increased sensitivity and provides greater precision in dose/response assessment, which will be of use in risk assessment, when the F1 animals are allowed to mature to pnd 95 than when they are sacrificed at pnd 21?

The RTI study (RTI 2003) cannot answer the question posed regarding sensitivity because it was not designed to do so. To address this question, a side-by-side comparison is needed between the current multigeneration protocol vs. an otherwise identical protocol except for pup retention. Because of design limitations, the results of the RTI study should not be used as justification for

modifying the design of the multigeneration study. The recently completed RTI study was designed only to focus on hazard identification issues, but lacked the complete dose-response necessary to address whether or not F1 retention to adulthood enhanced sensitivity. Also, it was not designed to determine the number of pups needed to be retained, the optimal age for retention, optimal housing of retaining offspring (a considerable logistical concern) and whether or not it is necessary to retain females. Until such data have been collected, we believe that it is premature to suggest multigeneration studies include extension of F1 animals.

The RTI study does show, however, that pnd 95 is **not** always more sensitive than pnd 21. With respect to comparison of results obtained from evaluation at pnd 21 vs. pnd 95, the following tables of data/information extracted from the RTI study (2003) show that for some measurements/observations/endpoints evaluation at pnd 95 provides some degree of additional sensitivity. However, this should not be over interpreted.

In Table 2, the results indicate that for 'gross observations of missing tissues' **all effects observed at pnd 95 were also observed at pnd 21.**

Table 2. Gross Observation of Missing Tissues (data reported in the RTI study)

| Tissue | Vinclozolin | | Dibutylphthalate | |
|------------------|-------------|-------------|------------------|-------------|
| | 50 mg/kg/d | 100 mg/kg/d | 100 mg/kg/d | 500 mg/kg/d |
| Epididymis | | | | |
| Pnd 21 | 0 (0.0) | 2 (3.1) | 0 (0.0) | 14 (21.5) |
| Pnd 95 | 0 (0.0) | 4 (5.4) | 0 (0.0) | 33 (44.6) |
| Prostate dorsal | | | | |
| Pnd 21 | 0 (0.0) | 21 (32.3) | 0 (0.0) | 1 (1.5) |
| Pnd 95 | 0 (0.0) | 17 (23.0) | 0 (0.0) | 3 (4.0) |
| Prostate ventral | | | | |
| Pnd 21 | 0 (0.0) | 5 (7.7) | 0 (0.0) | 2 (3.1) |
| Pnd 95 | 0 (0.0) | 12 (16.2) | 0 (0.0) | 3 (4.0) |

The tables below (Tables 3-6) indicate that some effects on structures -- those that increase in size at puberty -- were more easily detected in the pnd 95 group compared to the pnd 21 group. However, this information cannot be viewed in isolation, and needs to be evaluated in terms of the relevance for risk assessment and establishment of NOAELs and LOAELs. In this regard it is particularly important to recognize that RTI concluded (page 59) that *“adverse reproductive system effects in toto (structural malformations and other abnormalities) of the low and high doses of VIN [vinclozolin] and the high dose of DBP [di-n-butyl phthalate] on F1 adult male offspring would most likely be statistically significant with either one or three adult males/litter and would have been detected with either study design [pnd 21 or pnd 95].”*

Table 3. Gross Observation of Reduced Size (these structures exhibit pubertal sensitive size) (data reported in the RTI study) (shaded cells – effects observed at pnd 95 which were not observed at pnd 21)

| Tissue | Vinclozolin | | Dibutylphthalate | |
|------------------|-------------|-------------|------------------|-------------|
| | 50 mg/kg/d | 100 mg/kg/d | 100 mg/kg/d | 500 mg/kg/d |
| Epididymis | | | | |
| Pnd 21 | 0 (0.0) | 0 (0.0) | 0 (0.0) | 4 (6.2) |
| Pnd 95 | 0 (0.0) | 19 (25.7) | 0 (0.0) | 52 (71.6) |
| Prostate dorsal | | | | |
| Pnd 21 | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Pnd 95 | 3 (3.2) | 20 (27.0) | 2 (2.5) | 8 (10.8) |
| Prostate ventral | | | | |
| Pnd 21 | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Pnd 95 | 7 (7.4) | 45 (60.8) | 2 (2.5) | 8 (10.8) |

Table 4. Tissues Reported as Missing/Size Reduced (the RTI report did not distinguish between missing & reduced size of these tissues) (data reported in the RTI study) (shaded cells – effects observed at pnd 95 which were not observed at pnd 21)

| Tissue | Vinclozolin | | Dibutylphthalate | |
|------------------|-------------|-------------|------------------|-------------|
| | 50 mg/kg/d | 100 mg/kg/d | 100 mg/kg/d | 500 mg/kg/d |
| Cowper's glands | | | | |
| Pnd 21 | 9 (11.0) | 47 (72.3) | 1 (1.4) | 7 (10.8) |
| Pnd 95 | 7 (7.4) | 63 (85.1) | 0 (0.0) | 8 (10.8) |
| LABC | | | | |
| Pnd 21 | 0 (0.0) | 1 (1.5) | 0 (0.0) | 0 (0.0) |
| Pnd 95 | 2 (2.1) | 40 (54.0) | 0 (0.0) | 4 (5.4) |
| Seminal vesicles | | | | |
| Pnd 21 | 0 (0.0) | 8 (12.3) | 0 (0.0) | 7 (10.8) |
| Pnd 95 | 5 (6.1) | 63 (85.1) | 0 (0.0) | 39 (52.7) |

Table 5. Penile deformities (data reported in the RTI study) (shaded cells – effects observed at pnd 95 which were not observed at pnd 21)

| Tissue | Vinclozolin | | Dibutylphthalate | |
|-------------|-------------|-------------|------------------|-------------|
| | 50 mg/kg/d | 100 mg/kg/d | 100 mg/kg/d | 500 mg/kg/d |
| Epispadias | | | | |
| Pnd 21 | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Pnd 95 | 4 (4.3) | 11 (14.9) | 0 (0.0) | 0 (0.0) |
| Hypospadias | | | | |
| Pnd 21 | 8 (9.7) | 52 (80.0) | 0 (0.0) | 2 (3.1) |
| Pnd 95 | 15 (15.8) | 73 (98.6) | 0 (0.0) | 12 (16.2) |
| Cleft | | | | |
| Pnd 21 | 4 (4.9) | 25 (38.5) | 0 (0.0) | 2 (3.1) |
| Pnd 95 | 41 (43.2) | 74 (100.0) | 2 (2.5) | 26 (35.1) |

Table 6. Observations reported for of Testes (data reported in the RTI study)
(shaded cells – effects observed at pnd 95 which were not observed at pnd 21)

| Tissue | Vinclozolin | | Dibutylphthalate | |
|-----------------|-------------|-------------|------------------|-------------|
| | 50 mg/kg/d | 100 mg/kg/d | 100 mg/kg/d | 500 mg/kg/d |
| Undescended | | | | |
| Pnd 21 | 0 (0.0) | 3 (4.6) | 0 (0.0) | 4 (6.1) |
| Pnd 95 | 1 (1.0) | 15 (20.3) | 0 (0.0) | 10 (13.3) |
| Reduced in size | | | | |
| Pnd 21 | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Pnd 95 | 1 (1.0) | 17 (23.0) | 0 (0.0) | 45 (60.8) |

Table 7 summarizes the results of changes in tissue weights relative to body weights for the pnd 21 and pnd 95 groups. **Here the results show that there is not a consistent pattern, and that that pnd 95 is not always more 'sensitive' than pnd 21.** In fact, 10 'endpoints' were detected at pnd 21 only (& not at pnd 95), while 5 'endpoints' were detected at pnd 95 only (& not at pnd 21).

Table 7. Significant Tissue Weights – relative to body weight (data reported in the RTI study)
(shaded cells -- effects observed at pnd 95 which were not observed at pnd 21; cross hatched cells – effects observed at pnd 21 that were not observed at pnd 95)

| Tissue | Vinclozolin | | Dibutylphthalate | |
|-----------------------------|-------------|-------------|------------------|-------------|
| | 50 mg/kg/d | 100 mg/kg/d | 100 mg/kg/d | 500 mg/kg/d |
| Testes | | | | |
| L pnd 21 | | . | X | X |
| R pnd 21 | | . | X | X |
| L pnd 95 | | X | | X |
| R pnd 95 | | X | | X |
| Corpus and Caput Epididymis | | | | |
| L pnd 21 | | X | X | X |
| R pnd 21 | X | X | X | X |
| L pnd 95 | | X | | X |
| R pnd 95 | | X | | X |
| Cauda Epidid. | | | | |
| L pnd 21 | | X | | X |
| R pnd 21 | | X | | X |
| L pnd 95 | | X | | X |
| R pnd 95 | X | X | X | X |
| Seminal vesicles | | | | |
| Pnd 21 | X | X | | X |
| Pnd 95 | | X | | X |
| Whole Prostate | | | | |
| Pnd 21 | X | X | | X |
| Pnd 95 | | X | | X |
| Ventral Prostate | | | | |
| Pnd 21 | X | X | | X |
| Pnd 95 | | X | | X |
| Dorsolateral | | | | |
| Pnd 21 | | X | | X |
| Pnd 95 | | X | | X |
| LABC | | | | |
| Pnd 21 | X | X | | X |
| Pnd 95 | X | X | | X |
| Cowper's Glands | | | | |
| Pnd 21 | | | X | X |
| Pnd 95 | | X | | |

Comments on Continued Development of Information by EPA and the proposal for discussion by the EDMVS that until such information is available, EPA may encourage the optional extension of one or more additional F1 animals per sex per dose to adulthood in all cases where Tier I is bypassed, and of one or more additional F1 animals per dose of the appropriate sex(es) where Tier I information indicates interaction of the test chemical with the estrogen, androgen, and/or thyroid systems.

We acknowledge that for some types of testing (e.g., pharmaceuticals), tailoring certain protocols based on mechanistic information can be very useful. However, the proposal to extend additional F1 animals in the 2-generation study based on Tier I information is likely to present some problems that need to be considered. First, the 2-generation study guideline is used to evaluate a very wide range of chemicals that are used in a variety of ways that can change over time. Similarly, our knowledge of mechanisms of action can also change over time. This presents the potential problem of running a standard guideline 2-generation study (i.e., without the F1 extension), only to subsequently learn of new information suggesting hormonal activity. How would the adequacy of that 2-generation study be viewed? This scenario also applies to all of the existing 2-generation studies that have been conducted over its many years of use. This situation could result in many 2-generation studies having to be repeated with the value such repeats questionable. This is not a trivial matter considering the high costs and number of animals used in the study (>3000 per compound). It would seem far better to have one globally accepted protocol for a wide range of chemicals evaluated under a many different regulatory agency programs. Along these lines, it should be recognized that global regulatory agencies worked for several years to internationally harmonize the 2-generation study protocol, a process that has been completed only recently. Obviously, these agencies would not have expended such effort if protocol harmonization were not beneficial. The concept of an alternate 2-generation study protocol triggered by Tier I data runs counter to these harmonization efforts.

In exploring the utility of maintaining additional F1 males and females until adulthood, we believe the discussion has been disproportionately focused on detection of hazards, but has overlooked the overall impact for risk assessment. In terms of addressing hazards, it is self-evident that certain alterations to the male reproductive tract will be more readily discerned at the adult life stage than as a weanling, simply due to the small size of certain organs or other aspects of normal developmental biology. Therefore, it should be no surprise that the F1 extension study reported certain male reproductive tract malformations at a higher incidence in the F1 males retained until adulthood than in weanlings. However, it is well known that morphological alterations induced by hormonal agents tend to occur as syndromes, rather than isolated findings. This was confirmed and clearly shown in the F1 extension study with DBP and vinclozolin, both of which were detected as producing adverse effects upon androgen-sensitive tissues on the basis of standard end points (including malformations seen at weanling necropsy). Rather than trying to comprehensively characterize all possible developmental hazards caused by a compound, regardless of the dose needed to cause such effects, it would seem more prudent to focus on those developmental end points which appear to be most sensitive, and hence, most critical for estimating risk.

Based on the above, we agree that more quantitative data comparing various protocol modifications would be useful. However, this research needs to take a holistic approach, considering the value and relative sensitivity of various end points and various types of malformations. The recently completed F1 extension study was designed only to focus on hazard identification issues, but lacked the complete dose-response necessary to address whether or not F1 retention to adulthood enhanced sensitivity. Also, it was not designed to determine the number of pups needed to be retained, the optimal age for retention, optimal housing of retaining offspring (a considerable logistical concern) and whether or not it is necessary to retain females. Until such data have been collected, we believe that it is premature to suggest multigeneration studies include extension of F1 animals. The ability of the current multi-generation study is quite adequate to detect reproductive effects of weak androgens, including male repro tract malformations, as shown by the F1 extension study as well as multigeneration rat reproduction studies of weak anti-androgenic substances. For example, male reproductive tract malformations were found in multigeneration studies on the anti-androgenic compound, linuron (summarized in the publicly available EPA Registration Eligibility Document). In fact, these multigeneration studies were conducted according to pre-1998 guidelines, so a current two-generation study would only be more effective at detecting such effects.

References

- Clark, R.L. (1998). Endpoints of reproductive system development. In: *An Evaluation and Interpretation of Reproductive Endpoints for Human Risk Assessment*, pp. 27-62. International Life Sciences Institute, Health and Environmental Science Institute, Washington, D.C.
- DaCosta, V.M., Moreira, D.G., and Rosenthal, D. (2001). Thyroid function and aging: gender-related differences. *J. Endocrinol.* **171**, 193-198.
- Döhler, K.-D., Wong, C.C. and von zur Mühlen, A. (1979). The rat as model for the study of drug effects on thyroid function: consideration of methodological problems. *Pharmacol. Ther.* **5**, 305-318.
- Gallavan, R.H., Jr., Holson, J.F., Stump, D.G., Knapp, J.F., and Reynolds, V.L. (1999). Interpreting the toxicologic significance of alterations in anogenital distance: potential for confounding effects of progeny body weights. *Reprod. Toxicol.* **13**, 383-390.
- ILSI (1998). *An Evaluation and Interpretation of Reproductive Endpoints for Human Risk Assessment*. International Life Sciences Institute, Health and Environmental Science Institute, Washington, D.C.
- RTI (2003) Report on the One-Generation Extension Study of Vinclozolin and Di-N-Butyl Phthalate Administered by Gavage on Gestational Day 6 to Postnatal Day 20 in CD (Sprague-Dawley) Rats, EDMVS web site:
<http://www.epa.gov/scipoly/oscpendo/meetings/2003/june/onegenextensionfinalreport5803.pdf>